

Supplementary Tables and Figures S1-S7:

Table S2. Primer sequences used for combinatorial allele-specific PCR targeting exon 2 and flanking intronic and exonic regions

Primer Name	Sequence	Genic Region
<i>RHCE_2083_F</i>	GGTGTCAAGGGGAGGGCTTC	Intron 1
<i>RHCE_12211_R</i>	CCAGCTGTGTCTCCGGAAATT	Exon 2
<i>RHD_2086_F</i>	GTCAGGGGAGGGAGGTGA	Intron 1
<i>RHD_12211_R</i>	CCAGCTGTGTCTCCGGAAACTC	Exon 2
<i>RHCE_12120_F</i>	CTTCCCCCTCCTCCTTCTCA	Intron 1
<i>RHCE_18338_R</i>	CTTCCCCAAGACAGCACCG	Exon 3
<i>RHD_12120_F</i>	CTTCCCCCTCGTCCTTCTCG	Intron 1
<i>RHD_18338_R</i>	CTTCCCCAAGACAGCATCCA	Exon 3

Table S3. Vector-specific primers used for Sanger validation of the exon 2 gene conversion event (*RHCE*D-CE(2)CE*)

Gene	Region	Strand	Name	Sequence (5' to 3')
<i>RHD, RHCE</i>	Intron 1	Sense	RHDCE_10054_Fwd	TGCCACCTCATCCTCCTAA
<i>RHD, RHCE</i>	Intron 1	Antisense	RHDCE_10231_Rev	GTGCCAGCCTAGACAATGT
<i>RHD, RHCE</i>	Intron 1	Sense	RHDCE_10483_Fwd	AATTGTCTACATGCTGGTTGC
<i>RHD, RHCE</i>	Intron 1	Antisense	RHDCE_10622_Rev	CAATAATCACCACACTGTTTCG
<i>RHD, RHCE</i>	Intron 1	Sense	RHDCE_10789_Fwd	GCACAGGCTCCAGCAGATTA
<i>RHD, RHCE</i>	Intron 1	Antisense	RHDCE_10893_Rev	GATCTGCCACCTCGGTTTC
<i>RHD, RHCE</i>	Intron 1	Sense	RHDCE_11199_Fwd	TGGTGACAGTAACAGCACCC
<i>RHD, RHCE</i>	Intron 1	Antisense	RHDCE_11291_Rev	ACCATGTGCCTACGCTGTTT
<i>RHD, RHCE</i>	Intron 1	Sense	RHDCE_11577_Fwd	GATGTTTCTGCCAGCCAGC
<i>RHD, RHCE</i>	Intron 1	Antisense	RHDCE_11596_Rev	GCTGGCTGGCAGAAAACATC
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_13062_Rev	GGGGGCTGAAAGTCCTTCT
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_13530_Fwd	GCCAGGAACTGTGCTGAGCAC
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_13804_Rev	GAACCACCATGCCAGCCTC
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_13963_Fwd	CCACAGTCCCAGCTACTTGG
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_14211_Rev	GGGTTCTCCGAGTTAGCCA
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_14568_Fwd	GGCACCTGTAATCCCAGCTA
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_14894_Fwd	AAGGTACACACCAGCCTTCAGCC
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_15184_Rev	GTCCAAGTCAAGTGGCACAAGC
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_15554_Fwd	CACCAGGTGGTCAGGTTGG
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_15864_Rev	CCTGAGCAACACAGTGAGGT
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_16220_Fwd	CTCACTTGCCTTACCGTGGC
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_16560_Rev	GGAAGGAGTCTTACAGGGACCAGG
<i>RHD, RHCE</i>	Intron 2	Sense	RHDCE_16955_Fwd	GCCTCAGCCTCGAGAGTGGC
<i>RHD, RHCE</i>	Intron 2	Antisense	RHDCE_17273_Rev	GAGCAGCTGCTCTGAGACTCC
<i>pMiniT</i>	Vector Insert	Sense	pMiniT_Fwd	ACCTGCCAACCAAGCGAGAAC
<i>pMiniT</i>	Vector Insert	Antisense	pMiniT_Rev	TCAGGGTTATTGTCTCATGAGCG

Table S6. SNV-based *RHD* and *RHCE* alleles detected in Asian and Native American samples

Allele Name in ISBT v2.0 ¹	cDNA ¹	Phenotype ¹	Allele No. ²
<i>RHD</i> *01N.16	711delC	D null	2
<i>RHD</i> *01N.20	941G>T	D null	1
<i>RHD</i> *01W.1	809T>G	Weak D Type 1	2
<i>RHD</i> *01W.2	1154 G>C	Weak D Type 2	3
<i>RHD</i> *01W.33	520G>A	Weak D Type 33	2
<i>RHD</i> *[602C>G; 667T>G] ⁴	602C>G; 667T>G	Weak D Type 40, DFV	1
<i>RHD</i> *01W.45	1195G>A	Weak D Type 45	2
<i>RHD</i> *01W.66	916G>A	Weak D Type 66	2
<i>RHD</i> *DEL1	1227G>A	DEL	6
<i>RHD</i> *05.04	697G>C	DV Type 4	1
<i>RHD</i> *08.01	667T>G	DFV	6
<i>RHD</i> *[667T>G; 1136C>T] ⁴	667T>G; 1136C>T	-	1
<i>RHD</i> *10.00 ⁵	1136C>T	DAU0	140
<i>RHD</i> *16.01	667G>T; 676G>C	DCS1	1
<i>RHD</i> *33	1073T>C	DWI	1
<i>RHCE</i> *01.01	48G>C	e weak	1188
<i>RHCE</i> *[48G>C; 676G>C] ⁴	48G>C; 676G>C	e weak, E	72
<i>RHCE</i> *01.06	254C>G	partial e	2
<i>RHCE</i> *01.07	48G>C; 667G>T	partial e partial c	3
<i>RHCE</i> *01.20.01	733C>G	partial e partial c	3
<i>RHCE</i> *01.20.02	48G>C; 733C>G	partial e partial c	4
<i>RHCE</i> *01.20.05	48G>C; 1006G>T	partial e	1
<i>RHCE</i> *01.21	341G>A	RH:48 (JAL+)	1
<i>RHCE</i> *[48G>C; 122A>G] ⁴	48G>C; 122A>G	e weak, partial C	9
<i>RHCE</i> *03	676G>C	E ³	507

¹Allele names, cDNA changes and Phenotypes are as designated by ISBT v2.0 110914.²The number of alleles present in 1135 Asian and Native American samples.³ISBT phenotype was modified to avoid assuming phase between C, c and E, e indicative variants.⁴Novel *RH* allele relative to ISBT v2.0 110914. ISBT v5.0 was cross-referenced for updated nomenclature of novel *RH* alleles. The “[]” and “;” follow HGVS conventions to denote variants were statistically phased as present on the same chromosome.⁵Genotype quality was variable at this position.

Table S7. Predicted loss of function variants in *RHD* and *RHCE* detected in Asian and Native American samples and not annotated in ISBT v2.0 110914

Gene	Position ¹	Ref	Alt	Variant Type	Consequence	Allele Freq. ²	No. Alleles ³
<i>RHD</i>	chr1:25611057	C	T	Splice region	c.149-7C>T	0.0003	2
<i>RHCE</i>	chr1:25718638	A	G	Splice region	c.487-6T>C	0.00004	1

¹Position corresponds to the genomic position in GRCh37.

²Allele frequency corresponds to the maximum allele frequency reported in gnomAD.

³Number of variant alleles detected in this study.

Table S10. SNV-based RHD Null, weak D and partial D alleles detected in African American Samples

Allele Name in ISBT v2.0	Allele Name in ISBT v5.0	cDNA	Phenotype ¹	Allele No. ²	Allele Frequency (%)
<i>RHD*01N.18</i>		807T>G	D null	3	0.087
<i>RHD*01N.20</i>		941G>T	D null	3	0.087
<i>RHD*04N.01</i>	<i>RHD*08N.01</i>	37- bp insertion; 609G>A; 654G>C; 667T>G; 674C>T; 807T>G	D null	109	3.178
<i>RHD*01W.1</i>		809T>G	Weak D Type 1	2	0.058
<i>RHD*[809T>G; 1136C>T]⁴</i>		809T>G; 1136C>T	Weak D Type 1	1	0.029
<i>RHD*01W.33</i>		520G>A	Weak D Type 33	1	0.029
<i>RHD*01W.40</i>		602C>G	Weak D Type 40	1	0.029
<i>RHD*[602C>G; 667T>G]³</i>		602C>G; 667T>G	-	3	0.087
<i>RHD*[28C>T; 916G>A]³</i>		28C>T; 916G>A	-	1	0.029
<i>RHD*01W.66</i>		916G>A	Weak D Type 66	12	0.350
<i>RHD*[916G>A; 186G>T; 410C>T; 455A>C]³</i>		916G>A; 186G>T; 410C>T; 455A>C	-	1	0.029
<i>RHD*02</i>		1061C>A	DII	2	0.058
<i>RHD*[1061C>A; 1136C>T]^{3;4}</i>		1061C>A; 1136C>T	-	1	0.029
<i>RHD*03.01</i>		186G>T; 410C>T; 455A>C; 602C>G; 667T>G	DIIIa	9	0.262
<i>RHD*[186G>T; 410C>T; 455A>C; 602C>G; 667T>G; 1136C>T]^{3;4}</i>		186G>T; 410C>T; 455A>C; 602C>G; 667T>G; 1136C>T	-	2	0.058
<i>RHD*[186G>T; 410C>T; 455A>C; 602C>G; 667T>G; 819G>A]³</i>	<i>RHD*03.01</i>	186G>T; 410C>T; 455A>C; 602C>G; 667T>G; 819G>A	DIII	40	1.166
<i>RHD*[RHD*03; 186G>T; 410C>T; 602C>G; 667T>G; 819G>A; 1136C>T]^{3;4}</i>		186G>T; 410C>T; 455A>C; 602C>G; 667T>G; 819G>A; 1136C>T	-	9	0.262
<i>RHD*03.04</i>		186G>T; 410C>T; 455A>C	DIII type 4	77	2.245
<i>RHD*[186G>; 410C>T; 455A>C; 835G>A; 1136C>T]^{3;4}</i>		186G>; 410C>T; 455A>C; 835G>A; 1136C>T	-	3	0.087
<i>RHD*[186G>; 410C>T; 455A>C; 667T>G 697G>C; 1136C>T]^{3;4}</i>		186G>; 410C>T; 455A>C; 667T>G 697G>C; 1136C>T	-	1	0.029
<i>RHD*[186G>T; 410C>T; 455A>C; 667T>G]³</i>	<i>RHD*03.09</i>	186G>T; 410C>T; 455A>C; 667T>G	DIII type 4, DFV	1	0.029
<i>RHD*03.06</i>		410C>T; 455A>C; 602C>G; 667T>G; 819G>A	DIII type 6	2	0.058
<i>RHD*04.02</i>		186G>T; 410C>T; 455A>C; 1048G>C	DIVa type 2	29	0.845

<i>RHD</i> *[186G>T; 410C>T; 455A>C; 1048G>C; 1136C>T] ^{3,4}	186G>T; 410C>T; 455A>C; 1048G>C; 1136C>T	-	7	0.204	
<i>RHD</i> *05.01	667T>G; 697G>C	DV type 1	1	0.029	
<i>RHD</i> *[186G>T; 410C>T; 455A>C; 667T>G; 697G>C] ³	186G>T; 410C>T; 455A>C; 667T>G; 697G>C	DIII, DV	1	0.029	
<i>RHD</i> *08.01	667T>G	DFV	1	0.029	
<i>RHD</i> *09.03	602 C>G; 667 T>G	DAR1	49	1.429	
	819G>A				
<i>RHD</i> *[602 C>G; 667 T>G 819G>A; 1136C>T] ^{3,4}	602 C>G; 667 T>G 819G>A; 1136C>T		5	0.146	
<i>RHD</i> *10.00 ⁴	1136C>T	DAU0	763	22.24	
<i>RHD</i> *10.03 ⁴	835G>A; 1136C>T	DAU3	66	1.924	
<i>RHD</i> *10.04 ⁴	697G>A; 1136C>T	DAU4	3	0.087	
<i>RHD</i> *10.05 ⁴	667T>G; 697G>A; 1136C>T	DAU5	37	1.078	
<i>RHD</i> *12.01	509T>C; 667T>G	DOL1	5	0.146	
<i>RHD</i> *[509T>C; 667T>G; 1136C>T]	509T>C; 667T>G; 1136C>T	-	2	0.058	
<i>RHD</i> *12.02	509T>C; 667T>G; 1132C>G	DOL2	3	0.087	
<i>RHD</i> *[186G>T; 410C>T; 455A>C; 509T>C; 667T>G] ³	<i>RHD</i> *40	186G>T; 410C>T; 455A>C; 509T>C; 667T>G	-	1	0.029
<i>RHD</i> *17.04	505A>C; 509T>G	DFR4	1	0.029	
<i>RHD</i> *[938C>T; 1136C>T] ^{3,4}	938C>T; 1136C>T	-	1	0.029	
<i>RHD</i> *37	733G>C	DUC2	1	0.029	
<i>RHD</i> *weak D 4.2.2	602C>G; 667T>G; 744C>T; 1025T>C	DAR	11	0.321	
<i>RHD</i> *[602C>G; 667T>G; 697G>C; 744C>T; 1025 T>C] ³	602C>G; 667T>G; 697G>C; 744C>T; 1025 T>C	-	3	0.087	
<i>RHD</i> *[602C>G; 667T>G; 697G>C; 744C>T; 1025 T>C; 1136C>T] ^{3,4}	602C>G; 667T>G; 697G>C; 744C>T; 1025 T>C; 1136C>T	-	2	0.058	

¹Allele names, cDNA changes and Phenotypes are as designated by ISBT v2.0 110914.

²The number of alleles present in this dataset.

³Novel *RH* allele relative to ISBT v2.0 110914. The “[]” and “;” follow HGVS conventions to denote variants were statistically phased as present on the same chromosome.

⁴Genotype quality for the primary variant of the DAU cluster (NM_016124.3:c.1136C>T) was variable due to low coverage and sequence homology.

Table S11. SNV-based *RHCE* alleles detected in African American Samples

Allele Name in ISBT v2.0	cDNA ¹	Phenotype ¹	Allele No. ²	Allele Frequency (%)
<i>RHCE*01.01</i>	48G>C	e weak	1254	36.560
<i>RHCE*/[48G>C; 122A>G]⁴</i>	48G>C; 122A>G	-	6	0.175
<i>RHCE*/[48G>C; 106G>A]⁴</i>	48G>C; 106G>A	-	1	0.029
<i>RHCE*/[48G>C; 676G>C]⁴</i>	48G>C; 676G>C	-	15	0.437
<i>RHCE*/[48G>C; 602G>C]⁴</i>	48G>C; 602G>C	-	5	0.146
<i>RHCE*01.02</i>	48G>C; 1025C>T	partial e	67	1.953
<i>RHCE*01.03</i>	1025C>T	partial e	5	0.146
<i>RHCE*/[1025C>T; 733C>G]⁴</i>	1025C>T; 733C>G	-	1	0.029
<i>RHCE*01.04</i>	48G>C; 712A>G; 733C>G; 787A>G; 800T>A; 916A>G	partial e partial c	13	0.379
<i>RHCE*01.05</i>	48G>C; 712A>G; 787A>G; 800T>A	partial e partial c	11	0.321
<i>RHCE*/[48G>C; 712A>G; 787A>G; 800T>A; 733C>G]⁴</i>	712A>G; 787A>G; 800T>A; 733C>G	-	1	0.029
<i>RHCE*01.06</i>	254C>G	partial e	147	4.286
<i>RHCE*/[48G>C; 254C>G]⁴</i>	48G>C; 254C>G	-	15	0.437
<i>RHCE*/[254C>G; 872C>T]⁴</i>	254C>G; 872C>T	-	1	0.029
<i>RHCE*01.07</i>	48G>C; 667G>T	partial e partial c	41	1.195
<i>RHCE*01.08</i>	48G>C; 712A>G; 818C>T; 1132C>G	e+/-	10	0.292
<i>RHCE*01.20.01</i>	733C>G	partial e partial c	473	13.790
<i>RHCE*01.20.02</i>	48G>C; 733C>G	partial e partial c	119	3.469
<i>RHCE*01.20.03</i>	48G>C; 733C>G; 1006G>T	partial e partial c	122	3.557
<i>RHCE*01.20.04</i>	48G>C; 733C>G; 1025C>T	partial e	3	0.087
<i>RHCE*01.20.05</i>	733C>G; 1006G>T	partial e	9	0.262
<i>RHCE*01.20.06</i>	48G>C; 697C>G; 733C>G	partial e partial c	7	0.204
<i>RHCE*01.20.07</i>	340C>T; 733C>G	partial e	3	0.087
<i>RHCE*/[800T>A; 733C>G]⁴</i>	800T>A; 733C>G	-	1	0.029
<i>RHCE*03</i>	676G>C	E ³	363	10.583

¹Allele names, cDNA changes and Phenotypes are as designated by ISBT v2.0 110914.²The number of alleles present in this dataset.³The ISBT phenotype was modified to avoid assuming phase between C, c and E, e indicative variants.⁴Novel *RH* allele relative to ISBT v2.0 110914. ISBT v5.0 was cross-referenced for updated nomenclature of novel alleles. The “[]” and “;” follow HGVS conventions to denote variants were statistically phased as present on the same chromosome.

Table S12. Predicted loss of function variants in *RHD* and *RHCE* detected in African American samples and not annotated in ISBT v2.0 110914

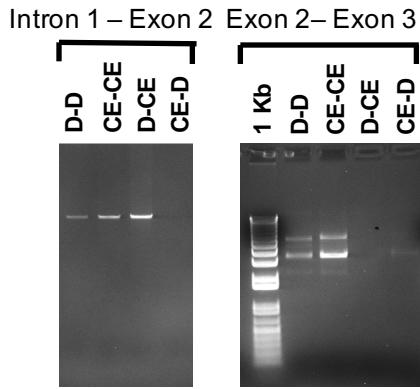
Gene	Position ¹	Ref	Alt	Variant Type	Consequence	Allele Freq. ²	No. Alleles ³
<i>RHD</i>	chr1:25611243	CTG	C	Frameshift	p.Phe111GlnfsTer48	0.00006	1
<i>RHD</i>	chr1:25633203	C	T	Frameshift	p.Gly353TrpfsTer39	0.0001	3
<i>RHD</i>	chr1:25628180	C	T	Splice region	c.801+3G>A	0.000009	1
<i>RHCE</i>	chr1:25729084	C	T	Splice region	c.486+3G>A	0.00004	1
<i>RHCE</i>	chr1:25735288	A	G	Stop gained	p.Trp74Ter	0.00001	1

¹Position corresponds to the genomic position in GRCh37. Frameshift variant positions are left-aligned.

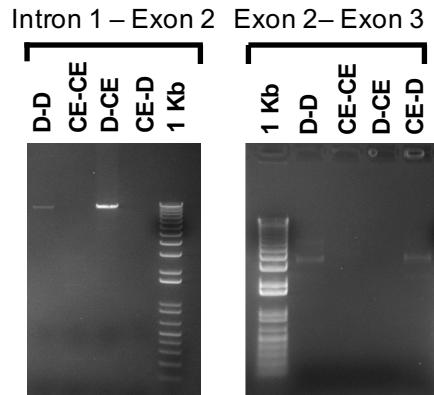
²Allele frequency corresponds to the maximum allele frequency reported in gnomAD

³Number of variant alleles detected in this study

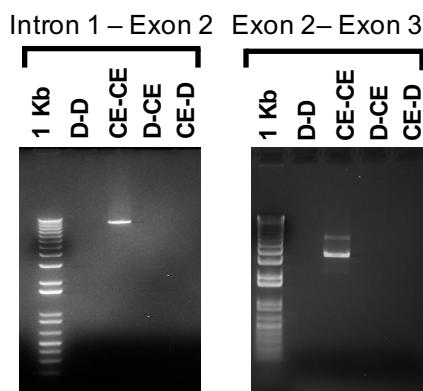
RBC1



RBC4



RBC5



RBC12

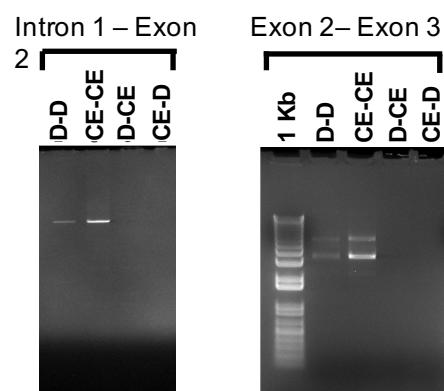


Figure S1. Combinatorial allele-specific PCR results for WHO reference samples. The results of combinatorial allele-specific PCR reactions are shown for each sample. PCR primer pairs were designed to target sequences between intron 1 and exon 2 and exon 2 and exon 3. *RHD* and *RHCE* are abbreviated D and CE, respectively. PCR products were visualized by gel electrophoresis on a 0.8% agarose gel (15 µl of PCR/lane).

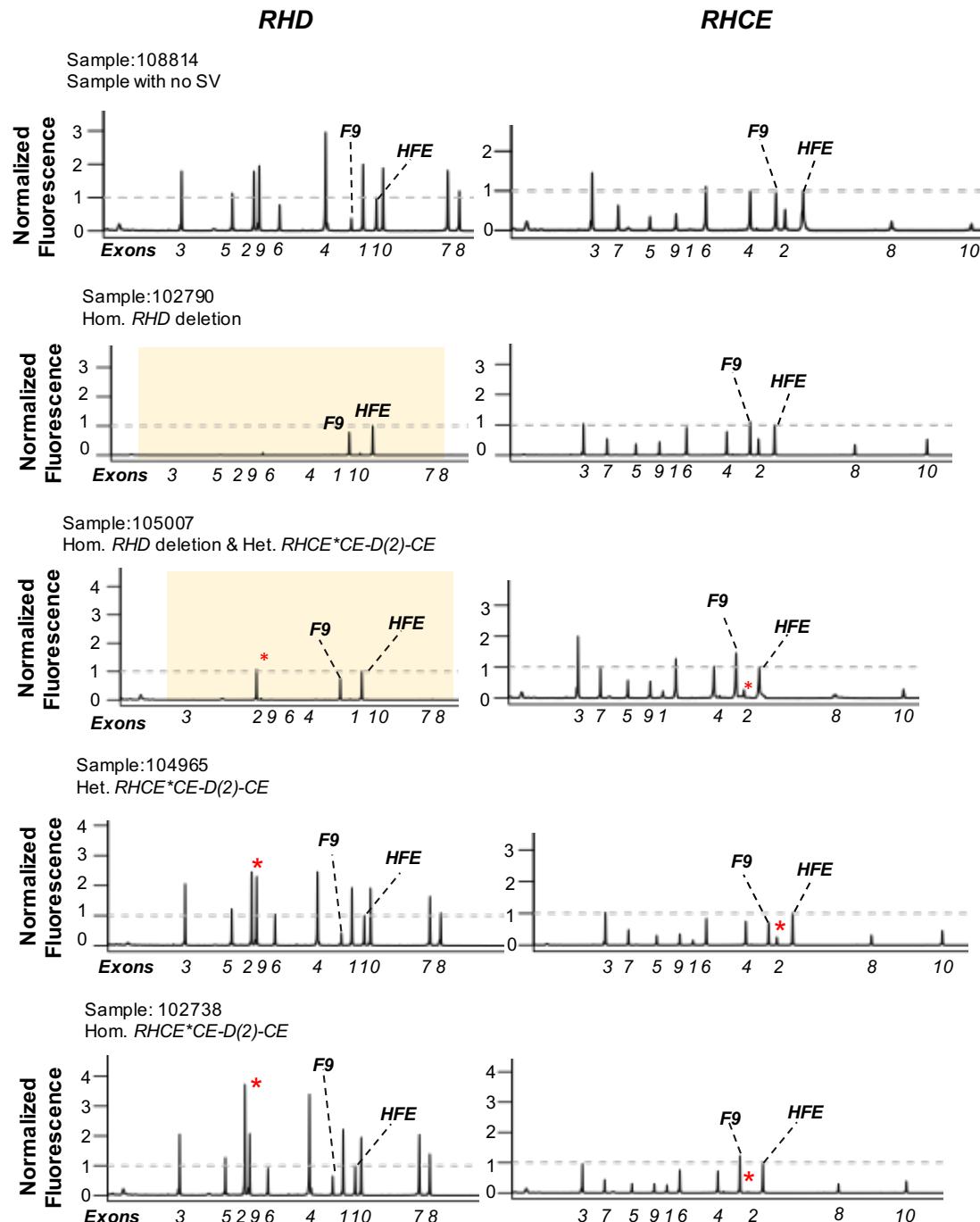


Figure S2. QMPSF results for Asian and Native American samples exhibiting no structural variation, *RHD* whole deletions and *RHCE*CE-D(2)-CE* hybrid alleles. Sample numbers and NGS-based structural variation events are shown with each row of panels showing the corresponding QMPSF results. Left panels show QMPSF results *RHD* and right panels show results for *RHCE*. The QMPSF peak heights are fluorescence measurements normalized to the amplified *HFE* exon. The exon amplified for *F9* is an additional positive amplification control. Light yellow panels highlight *RHD* whole gene deletions. (*) highlight individual exons with results indicative of structural variation.

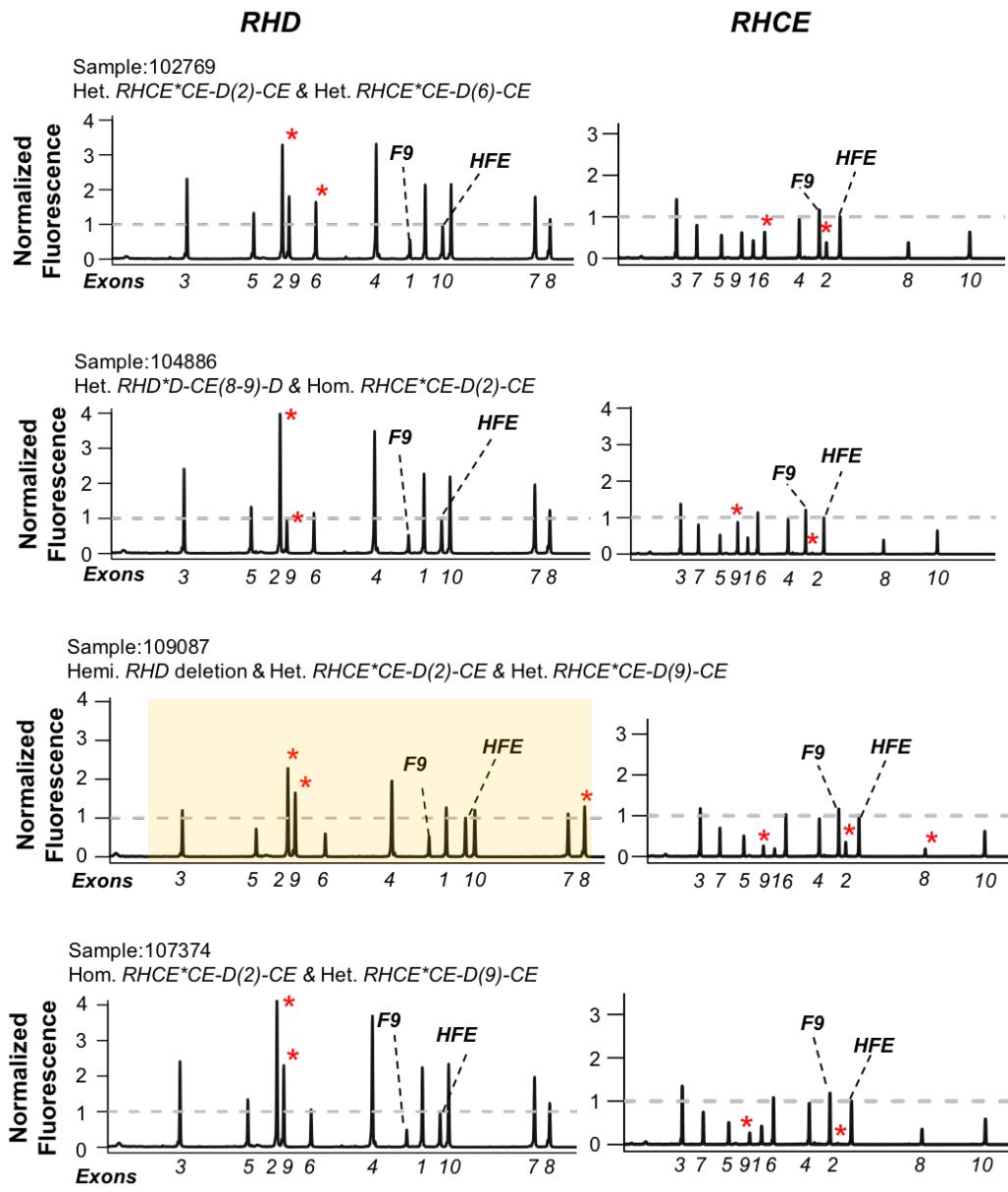


Figure S3. QMPSF results for Asian and Native American samples exhibiting *RHCE*CE-D(2)-CE* hybrid alleles and samples exhibiting structural variation in exons 6, 8, 9. Sample numbers and NGS-based structural variation events are labeled above each row of panels with the corresponding QMPSF results shown below. Left panels show QMPSF results for *RHD* and right panels show results for *RHCE*. The QMPSF peak heights are fluorescence measurements normalized to the amplified *HFE* exon. The exon amplified for *F9* serves as an additional positive amplification control. (*) highlight individual exons indicative of structural variation. Note for sample 104886, sample 109087 and samples 107374 there are discrepancies between QMPSF and NGS-based predictions with regards to structural variation affecting exon 8.

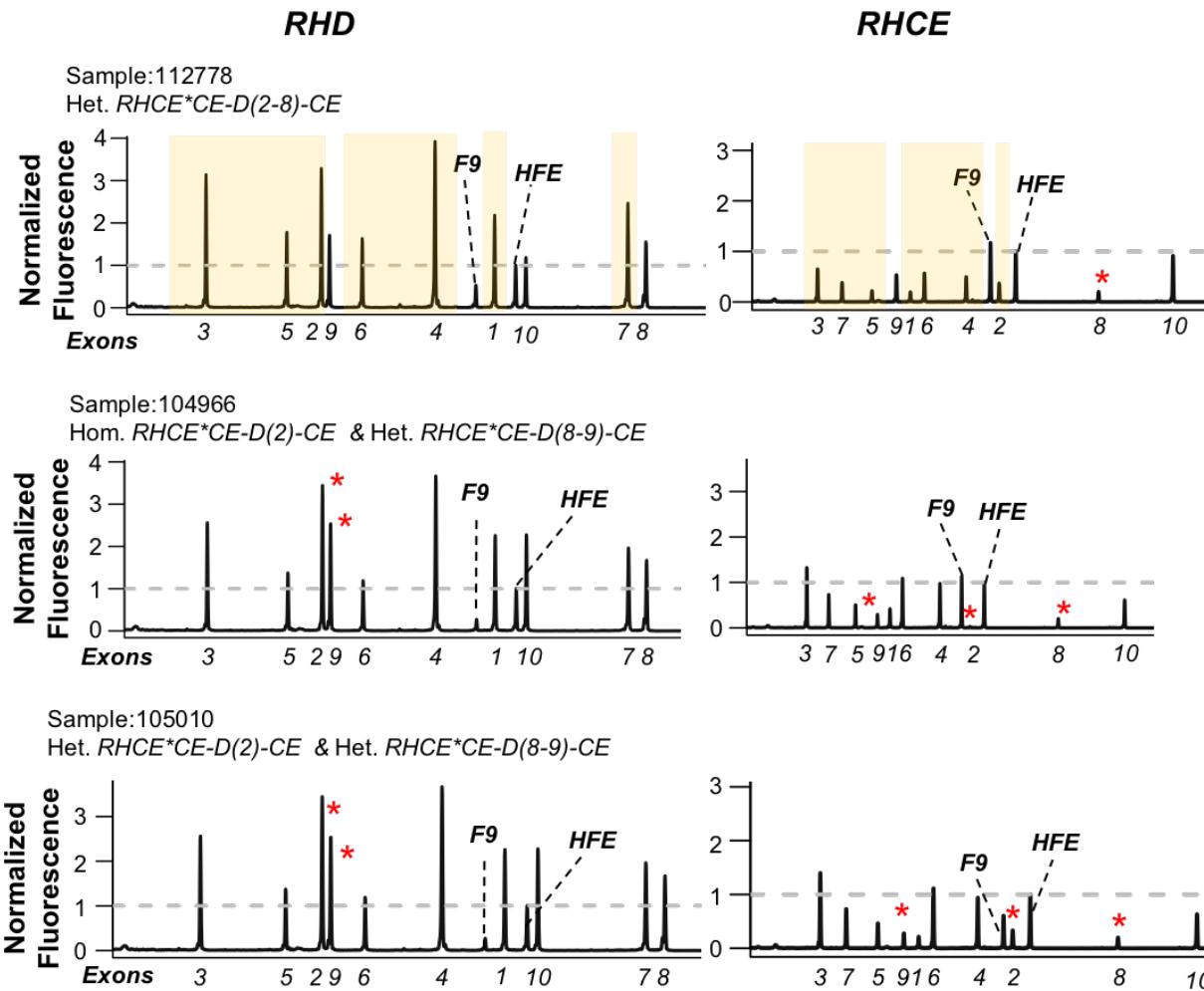


Figure S4. QMPSF results for Asian and Native American samples exhibiting *RHCE*CE-D(2)-CE* hybrid alleles and Asian and Native American samples exhibiting structural variation in exons 2-8 and 8-9.** Sample numbers and NGS-based structural variation events are shown above each row of panels with the corresponding QMPSF results shown below. Left panels show QMPSF results for *RHD* and right panels show results for *RHCE*. The QMPSF peak heights are fluorescence measurements normalized to the amplified exon in *HFE*. The amplified *F9* exon is an additional positive amplification control. Light yellow panels highlight structural variation events impacting multiple exons. (*) highlight amplicons with results indicative of structural variation.

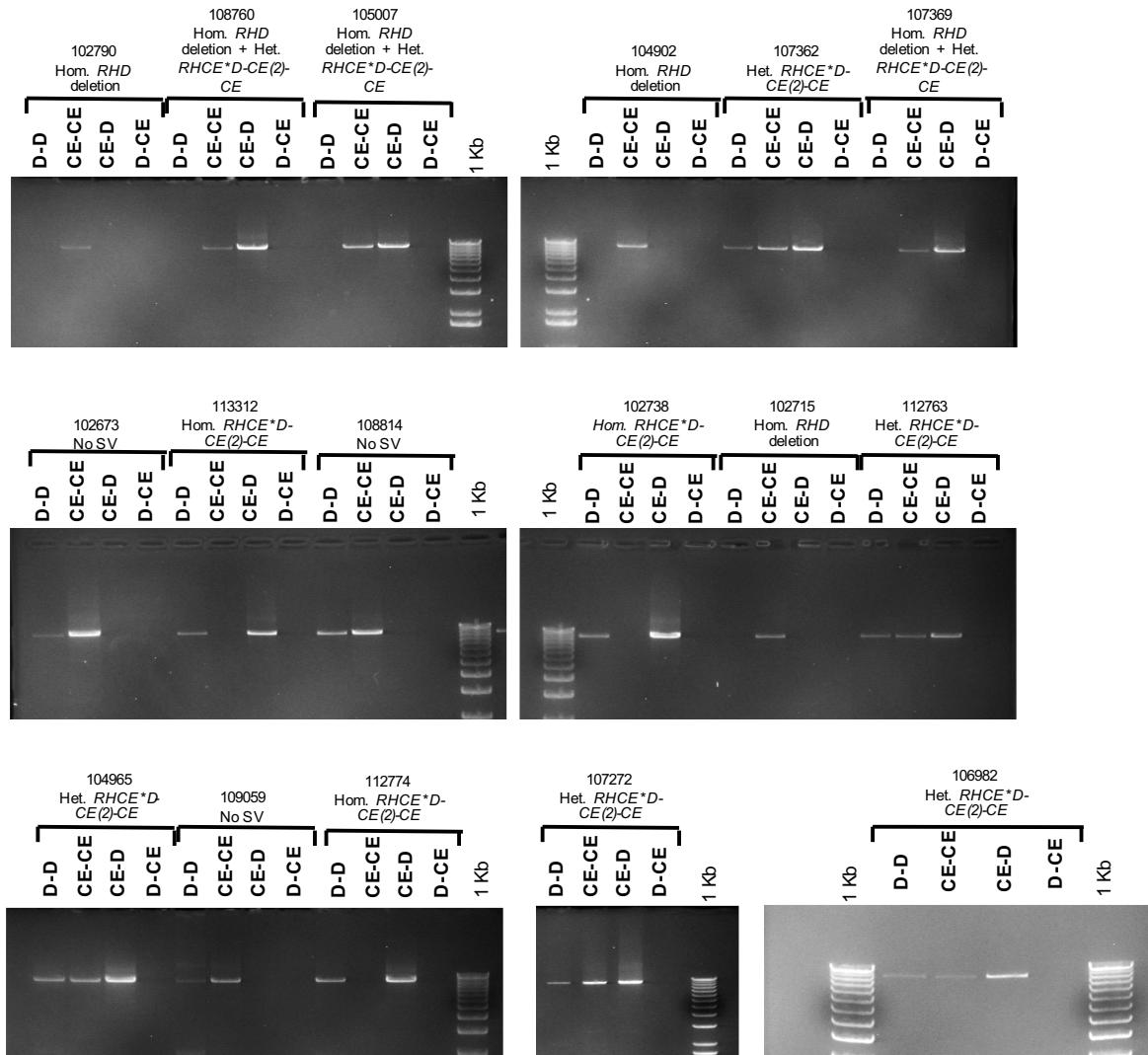


Figure S5. Combinatorial allele-specific PCR results (intron 1 – exon 2) for Asian and Native American samples exhibiting no structural variation, *RHD* whole deletions and *RHCE-CE-D(2)-CE* alleles. The PCR reactions shown correspond to primer pairs amplifying sequences between intron 1 – exon 2. The sample names are shown above corresponding PCR reactions with the NGS-predicted structural variation event. *RHD* and *RHCE* primers are abbreviated D and CE, respectively. PCR products were visualized by gel electrophoresis, loading 15 μ l of each PCR reaction/lane onto a 0.8% agarose gel.

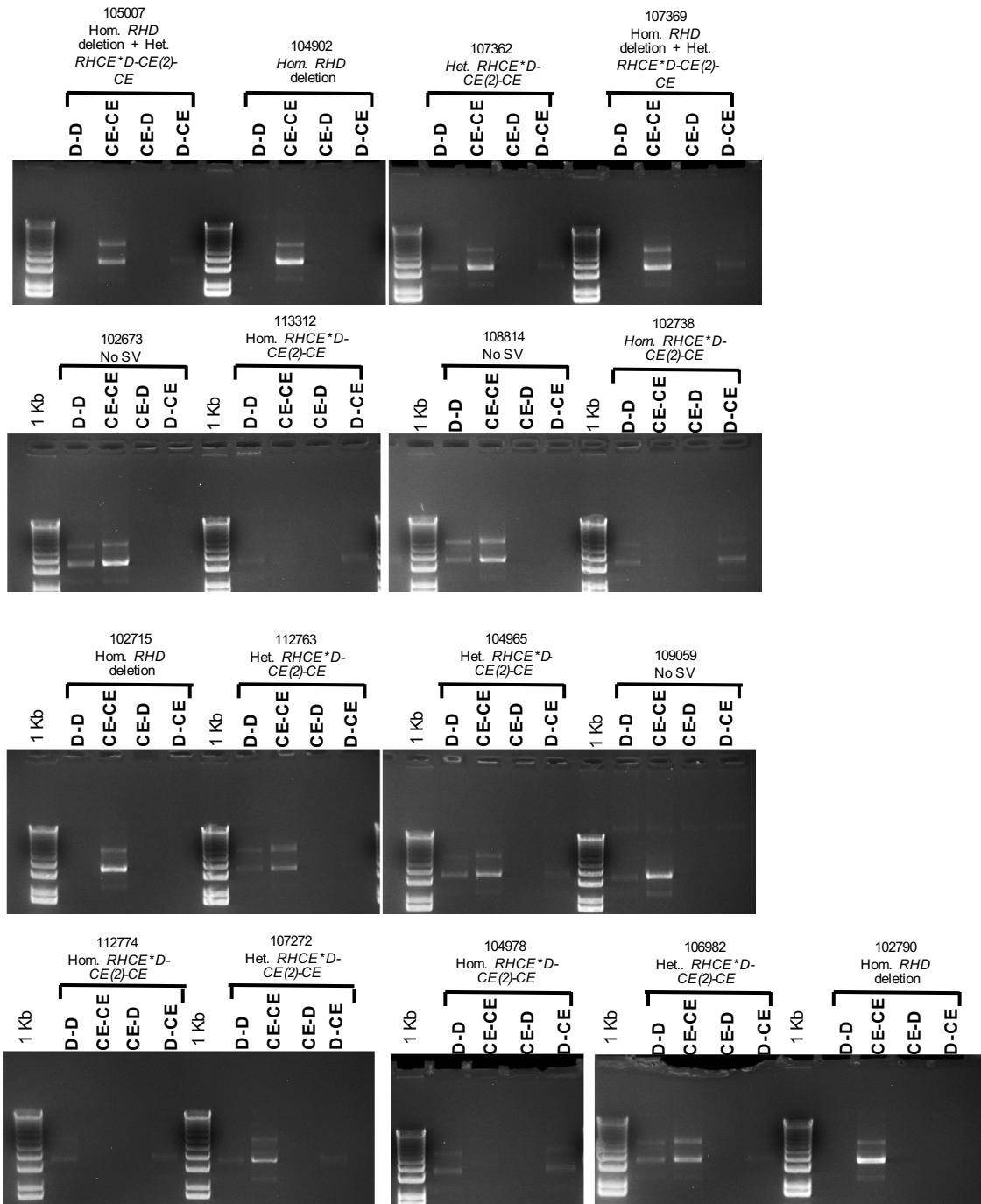
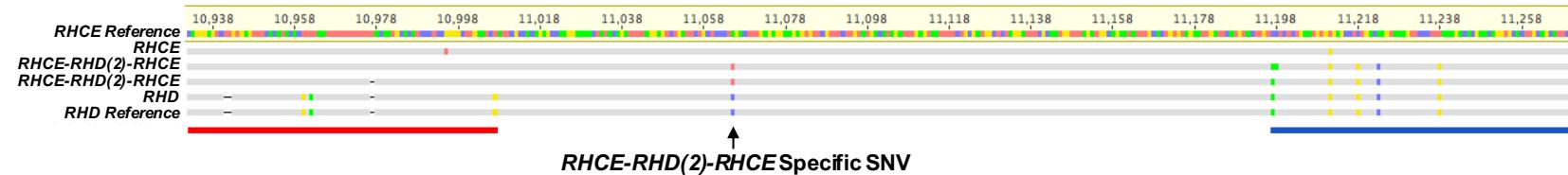


Figure S6. Combinatorial allele-specific PCR results (exon 2 – exon 3) for Asian and Native American samples exhibiting no structural variation, *RHD* whole deletions and *RHCE-CE-D(2)-CE* alleles. The PCR reactions shown correspond to primer pairs amplifying sequences between exon 2 – exon 3. The sample names are shown above corresponding PCR reactions with the NGS-predicted structural variation event. *RHD* and *RHCE* primers are abbreviated D and CE, respectively. PCR products were visualized by gel electrophoresis, loading 15 µl of each PCR reaction/lane onto a 0.8% agarose gel.

A) RHCE-RHD(2)-RHCE Intron 1 Upstream Breakpoint Region



B) RHCE-RHD(2)-RHCE Intron 2 Downstream Breakpoint Region

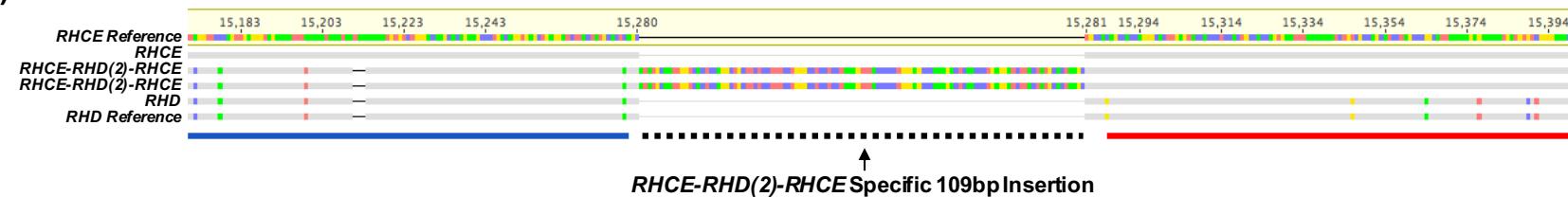


Figure S7: Validation of the common *RHCED-CE(2)CE intronic breakpoints by Sanger sequencing of cloned intron 1 (upstream) and intron 2 (downstream) allele-specific PCR products from two *RHCE**D-CE(2)CE Asian and Native American samples and the corresponding *RHD* and *RHCE* regions from a non-hybrid sample.** DNA sequences were aligned against the reference sequence for *RHCE* (top) with the *RHD* reference sequence aligned against *RHCE* on the bottom. Colored rectangles represent base pairs variant relative to the *RHCE* reference sequence. A) In intron 1, *RHCE* variants are identified in the *RHCE**D-CE(2)CE hybrid alleles (region underscored in red) until a single intron 1 nucleotide at position 11,065 that is an A in the *RHCE* reference, a C in the *RHD* reference, and a T in the *RHCE**D-CE(2)CE hybrid alleles. After this position, the hybrid alleles have *RHD* variants (region underscored in blue). This SNV is common and specific for *RHCE**D-CE(2)CE alleles in our study. B) In intron 2, *RHCE**D-CE(2)CE hybrid alleles have *RHD* variants (region underscored in blue) until a 109bp region of novel sequence, after which the hybrid alleles have *RHCE* variants. This sequence is consistent with the previously reported *RHCE* 109bp intron 2 “insertion” and likely identifies the precise downstream breakpoint in these hybrid alleles.